

Pathwise Sensitivity Analysis in Transient Regimes

Georgios Arampatzis^{*1}, Markos A. Katsoulakis^{†1}, and Yannis Pantazis^{‡1}

¹Department of Mathematics and Statistics University of Massachusetts, Amherst, MA, USA

Abstract

The instantaneous relative entropy (IRE) and the corresponding instantaneous Fisher information matrix (IFIM) for transient stochastic processes are presented in this paper. These novel tools for sensitivity analysis of stochastic models serve as an extension of the well known relative entropy rate (RER) and the corresponding Fisher information matrix (FIM) that apply to stationary processes. Three cases are studied here, discrete-time Markov chains, continuous-time Markov chains and stochastic differential equations. A biological reaction network is presented as a demonstration numerical example.

1 Introduction

Sensitivity analysis, for a general mathematical model, is defined to be the quantification of system response to parameter perturbations. Questions on the robustness, (structural) identifiability, experimental design, uncertainty quantification, estimation and control can be addressed through sensitivity analysis [1]. Moreover, it is a necessary analysis tool for the study of kinetic models such as chemical and biochemical reaction networks [1, 2]. The mathematical models considered in this paper are models that describe phenomena exhibiting stochasticity: stochastic differential equations (SDE), discrete-time and continuous-time Markov chains (DTMC and CTMC). Some of the mathematical tools for the sensitivity analysis of such systems include log-likelihood methods and Girsanov transformations [3, 4, 5], polynomial chaos [6], finite difference methods [7, 8] and pathwise sensitivity methods [9]. Closely related to the log-likelihood methods are various linear-response-based approaches, for instance in the context of chemical kinetics [10], as well as in recent mathematical work for linear response in non-equilibrium systems [11, 12].

In [13], the authors propose a new methodology for the pathwise sensitivity analysis of complex stochastic stationary dynamics based on the Relative Entropy (RE) between path distributions and Relative Entropy Rate (RER). This two quantities provide a measure of the sensitivity of the entire time-series distribution. The space of all such time-series is referred in probability theory as the “path space”. RER measures the loss of information per unit time in path space after an arbitrary perturbation of parameter combinations. Moreover, RER and the corresponding Fisher Information Matrix (FIM) become computationally feasible in certain cases as they admit explicit formulas. In fact, it is been showed in [13] that

^{*}arampatzis@math.umass.edu

[†]markos@math.umass.edu

[‡]pantazis@math.umass.edu

the proposed pathwise sensitivity analysis has the following properties: (a) it is rigorously valid for the sensitivity of long-time, stationary dynamics, (b) it is a gradient-free sensitivity analysis method suitable for high-dimensional parameter spaces, (c) the computation of RER and FIM does not require the explicit knowledge of the equilibrium probability distribution, relying only on information for local dynamics, i.e. it is suitable for non-equilibrium systems.

In this paper we extend the RE and FIM tools, developed in [13] for stationary processes, to transient and non-stationary processes. The extension is based on the notion of instantaneous RE (IRE) and instantaneous FIM (IFIM), see for example equations (10) and (12) in text. These sensitivity tools provide an instantaneous measure for the sensitivity of a system arising from information theoretic tools. Moreover, both IRE and IFIM, which are independent of observable functions, can be used as upper bounds for observable depended sensitivities, see for example the discussion in Conclusions and [14, 15].

The rest of the paper is organized as follows: In Section 2 a review of the RE and the FIM is given. In Sections 3,4 and 5 the new quantities IRE and IFIM are presented for discrete-time Markov chains, continuous-time Markov chains and stochastic differential equations, respectively. In Section 6, a numerical example for a biological reaction network is utilized and IRE and IFIM are computed in the course of a stochastic simulation and various observations are discussed. Finally, in Section 7, concluding remarks and connections with existing works are discussed.

2 Time-dependent sensitivity analysis

2.1 Decomposition of the pathwise relative entropy

The relative entropy (or Kullback-Leibler divergence) of a probability measure P with respect to (w.r.t.) another probability measure \bar{P} is defined as [16, 17]

$$\mathcal{R}(P|\bar{P}) := \begin{cases} \int \log \frac{dP}{d\bar{P}} dP, & P \ll \bar{P} \\ \infty, & \text{otherwise} \end{cases} \quad (1)$$

where $\frac{dP}{d\bar{P}}$ is a function known as the Radon-Nikodym derivative which is well-defined when P is absolutely continuous w.r.t. \bar{P} (denoted as $P \ll \bar{P}$) while the integration is performed w.r.t. the probability measure P . Relative entropy has been utilized in a diverse range of scientific fields from statistical mechanics [18] to telecommunications [17] and finance [19] and possesses the following fundamental properties:

- (i) $\mathcal{R}(P|\bar{P}) \geq 0$,
- (ii) $\mathcal{R}(P|\bar{P}) = 0$ if and only if $P = \bar{P}$ P -almost everywhere, and,
- (iii) $\mathcal{R}(P|\bar{P}) < \infty$ if and only if P and \bar{P} are absolutely continuous w.r.t. each other.

These properties allow us to view relative entropy as a “distance” (more precisely a divergence) between two probability measures capturing the relative importance of uncertainties [20]. From an information theory perspective, relative entropy measures the *loss/change of information* when \bar{P} is considered instead of P [17].

Let a stochastic process –either discrete-time or continuous-time– be denoted by X_t and let the path space \mathcal{X} be the set of all trajectories $\{X_t\}_{t=0}^T$. We denote by $Q_{0:T}$ the path space distribution, i.e., the probability to see a particular element of path space, \mathcal{X} . Denote by $\bar{Q}_{0:T}$ the path space distribution of another process, \bar{X}_t . The pathwise relative entropy of the distribution $Q_{0:T}$ w.r.t. the distribution $\bar{Q}_{0:T}$ assuming that they are absolutely continuous w.r.t. each other is written using (1) as

$$\mathcal{R}(Q_{0:T}|\bar{Q}_{0:T}) = \int \log \frac{dQ_{0:T}}{d\bar{Q}_{0:T}} dQ_{0:T}. \quad (2)$$

A key property of pathwise relative entropy is that it is an increasing function of time which is the analog of the second thermodynamic law in statistical physics [17]. To this end, for various important Markov processes, it can be shown, exploiting the Markov property, that the pathwise relative entropy can be written as an averaged quantity as

$$\mathcal{R}(Q_{0:T} | \bar{Q}_{0:T}) = \begin{cases} \mathcal{R}(\nu | \bar{\nu}) + \sum_{i=1}^T \mathcal{H}(Q_i | \bar{Q}_i) , & \text{if time is discrete} \\ \mathcal{R}(\nu | \bar{\nu}) + \int_0^T \mathcal{H}(Q_t | \bar{Q}_t) dt , & \text{if time is continuous ,} \end{cases} \quad (3)$$

where $\mathcal{R}(\nu | \bar{\nu})$ is the relative entropy of the initial distribution ν w.r.t. the perturbed initial distribution $\bar{\nu}$, while $\mathcal{H}(Q | \bar{Q})$ denotes the instantaneous relative entropy (for explicit formulas we refer to equation (10), (16) and (24)). We named the quantity $\mathcal{H}(Q | \bar{Q})$ instantaneous relative entropy because it is also non-negative but more importantly is the time-derivative of the pathwise relative entropy and inherits the properties of the relative entropy. Notice that the notation for instantaneous relative entropy is somewhat confusing but it will become clear when specific examples will be presented in the following sections.

Moreover the relative entropy rate is defined for a broad class of stochastic processes including Markov and semi-Markov processes as [21] as the limit

$$\mathcal{H}(Q | \bar{Q}) = \lim_{T \rightarrow \infty} \frac{1}{T} \mathcal{R}(Q_{0:T} | \bar{Q}_{0:T}) . \quad (4)$$

Relative entropy rate is closely related to the instantaneous relative entropy in two fundamental ways, Firstly it is the limit of instantaneous relative entropy as time goes to infinity, whenever this limit is defined and second it holds at the stationary regime that

$$\mathcal{H}(Q_i | \bar{Q}_i) = \mathcal{H}(Q | \bar{Q}) \quad (5)$$

for the discrete-time case with $i = 1, \dots, T$ and similarly for the continuous-time case.

2.2 Sensitivity analysis and Fisher information matrix

In [13], authors proposed the relative entropy between path distributions to perform sensitivity analysis arguing that pathwise relative entropy takes into account not only the equilibrium properties but also the complex dynamics of a stochastic process. Specifically, denote by $Q_{0:T}^\theta$ the path space distribution of the process X_t , parametrized by the parameter vector $\theta \in \mathbb{R}^K$. Consider also a perturbation vector, $\epsilon \in \mathbb{R}^K$, and denote by $Q_{0:T}^{\theta+\epsilon}$ the path space distribution of the perturbed process \tilde{X}_t . The pathwise relative entropy of the distribution $Q_{0:T}^\theta$ w.r.t. the distribution $Q_{0:T}^{\theta+\epsilon}$ assuming that they are absolutely continuous w.r.t. each other is written from (2) as

$$\mathcal{R}(Q_{0:T}^\theta | Q_{0:T}^{\theta+\epsilon}) = \int \log \frac{dQ_{0:T}^\theta}{dQ_{0:T}^{\theta+\epsilon}} dQ_{0:T}^\theta . \quad (6)$$

An attractive approach to sensitivity analysis that is rigorously based on relative entropy calculations is the Fisher Information Matrix (FIM). Indeed, assuming smoothness in the parameter vector, it is straightforward to obtain the following expansion for (6) [17, 22],

$$\mathcal{R}(Q_{0:T}^\theta | Q_{0:T}^{\theta+\epsilon}) = \frac{1}{2} \epsilon^T \mathcal{I}(Q_{0:T}^\theta) \epsilon + O(|\epsilon|^3) , \quad (7)$$

where the $K \times K$ pathwise FIM $\mathcal{I}(Q_{0:T}^\theta)$ is defined as the Hessian of the pathwise relative entropy. As (7) readily suggests, relative entropy is locally a quadratic function of the parameter vector θ . Indeed, the pathwise RE for any perturbation can be recovered up to third-order utilizing only the pathwise FIM. Moreover, similar expansions hold for the

instantaneous relative entropy and the relative entropy rate. The decomposition of the pathwise FIM reads

$$\mathcal{I}(Q_{0:T}^\theta) = \begin{cases} \mathcal{I}(\nu^\theta) + \sum_{i=1}^T \mathcal{I}_{\mathcal{H}}(Q_i^\theta), & \text{if time is discrete} \\ \mathcal{I}(\nu^\theta) + \int_0^T \mathcal{I}_{\mathcal{H}}(Q_t^\theta) dt, & \text{if time is continuous} \end{cases} \quad (8)$$

where $\mathcal{I}(\nu^\theta)$ is the FIM of the initial distribution ν^θ while $\mathcal{I}_{\mathcal{H}}(Q^\theta)$ denotes the instantaneous FIM. In the following sections concrete examples of stochastic Markov processes are presented whose pathwise relative entropy and the associated pathwise FIM is provided.

3 Discrete-time Markov chains

This section presents explicit formulas of the various relative entropy quantities defined in the previous section as well as the associated FIMs for the case of discrete-time Markov chains. The analysis of the DTMC case serves (a) as a more intuitive and manageable example of stochastic processes and (b) as a intermediate step to handle the continuous-time Markov chain case.

Next, let $\{x_i\}_{i \in \mathbb{Z}^+}$ be a discrete-time time-homogeneous Markov chain with separable state space E . The transition probability kernel of the Markov chain denoted by $P^\theta(x, dx')$ depends on the parameter vector $\theta \in \mathbb{R}^K$. Assume that the transition kernel is absolutely continuous with respect to the Lebesgue measure and the transition probability density function $p^\theta(x, x')$ is always positive for all $x, x' \in E$ and for all $\theta \in \mathbb{R}^K$. Exploiting the Markov property, the path space probability density $Q_{0:T}^\theta$ for the path $\{x_i\}_{i=0}^T$ at the time horizon $0, 1, \dots, T$ starting from the initial distribution $\nu^\theta(x)dx$ is given by

$$Q_{0:T}^\theta(x_0, \dots, x_T) = \nu^\theta(x_0)p^\theta(x_0, x_1) \dots p^\theta(x_{T-1}, x_T).$$

We consider a perturbation vector $\epsilon \in \mathbb{R}^K$ and the Markov chain $\{\bar{x}_i\}_{i \in \mathbb{Z}^+}$ with transition probability density function, $p^{\theta+\epsilon}(x, x')$, initial density, $\nu^{\theta+\epsilon}(x)$, as well as path distribution $Q_{0:T}^{\theta+\epsilon}$. The product representation of the path distributions results in an additive representation of the relative entropy of the path distribution $Q_{0:T}^\theta$ w.r.t. the perturbed path distribution $Q_{0:T}^{\theta+\epsilon}$. Let $\nu_i^\theta(x)$ denote the probability density function of the Markov chain at time instant i given that the initial distribution is ν^θ , whose formula is provided by the Chapman-Kolmogorov equation,

$$\nu_i^\theta(x) = \int_E \dots \int_E \nu^\theta(x_0)p^\theta(x_0, x_1) \dots p^\theta(x_{i-1}, x) dx_0 \dots dx_{i-1}.$$

The following theorem presents the decomposition of the pathwise relative entropy.

Theorem 3.1. (a) *The pathwise relative entropy for the above-defined discrete-time Markov chain is decomposed as*

$$\mathcal{R}(Q_{0:T}^\theta | Q_{0:T}^{\theta+\epsilon}) = \mathcal{R}(\nu^\theta | \nu^{\theta+\epsilon}) + \sum_{i=1}^T \mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon}), \quad (9)$$

where the instantaneous relative entropy equals to

$$\mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon}) = \mathbb{E}_{\nu_{i-1}^\theta} \left[\int_E p^\theta(x, x') \log \frac{p^\theta(x, x')}{p^{\theta+\epsilon}(x, x')} dx' \right] \quad (10)$$

(b) *Under smoothness assumption on the transition probability function for the parameter θ , the pathwise FIM is also decomposed as*

$$\mathcal{I}(Q_{0:T}^\theta) = \mathcal{I}(\nu^\theta) + \sum_{i=1}^T \mathcal{I}_{\mathcal{H}}(Q_i^\theta), \quad (11)$$

where the instantaneous pathwise FIM is given by

$$\mathcal{I}_{\mathcal{H}}(Q_i^\theta) = \mathbb{E}_{\nu_{i-1}^\theta} \left[\int_E p^\theta(x, x') \nabla_\theta \log p^\theta(x, x') \nabla_\theta \log p^\theta(x, x')^T dx' \right] \quad (12)$$

Proof. (a) The proof of this part of the theorem can be found in [17, Ch. 2] under the title “Chain rule for relative entropy” but for the shake of completeness we present it here. The Radon-Nikodym derivative of the unperturbed path distribution w.r.t. the perturbed path distribution takes the form

$$\frac{dQ_{0:T}^\theta}{dQ_{0:T}^{\theta+\epsilon}}(\{x_i\}_{i=0}^T) = \frac{\nu^\theta(x_0) \prod_{i=0}^{T-1} p^\theta(x_i, x_{i+1})}{\nu^{\theta+\epsilon}(x_0) \prod_{i=0}^{T-1} p^{\theta+\epsilon}(x_i, x_{i+1})},$$

which is well-defined since the transition probabilities are always positive. Then,

$$\begin{aligned} \mathcal{R}(Q_{0:T}^\theta | Q_{0:T}^{\theta+\epsilon}) &= \int_E \cdots \int_E \nu^\theta(x_0) \prod_{j=1}^T p^\theta(x_{j-1}, x_j) \log \frac{\nu^\theta(x_0) \prod_{i=1}^T p^\theta(x_{i-1}, x_i)}{\nu^{\theta+\epsilon}(x_0) \prod_{i=1}^T p^{\theta+\epsilon}(x_{i-1}, x_i)} dx_0 \dots dx_T \\ &= \int_E \cdots \int_E \nu^\theta(x_0) \prod_{j=1}^T p^\theta(x_{j-1}, x_j) \log \frac{\nu^\theta(x_0)}{\nu^{\theta+\epsilon}(x_0)} dx_0 \dots dx_T \\ &\quad + \sum_{i=1}^T \int_E \cdots \int_E \nu^\theta(x_0) \prod_{j=1}^T p^\theta(x_{j-1}, x_j) \log \frac{p^\theta(x_{i-1}, x_i)}{p^{\theta+\epsilon}(x_{i-1}, x_i)} dx_0 \dots dx_T \\ &= \mathcal{R}(\nu^\theta | \nu^{\theta+\epsilon}) + \sum_{i=1}^T \mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon}), \end{aligned}$$

where $\mathcal{R}(\nu^\theta | \nu^{\theta+\epsilon}) = \mathbb{E}_{\nu^\theta} \left[\log \frac{\nu^\theta(x)}{\nu^{\theta+\epsilon}(x)} \right]$ is the relative entropy of the unperturbed initial distribution w.r.t. the perturbed one, while the instantaneous relative entropy (of the time-varying pathwise relative entropy) is

$$\mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon}) = \mathbb{E}_{\nu_{i-1}^\theta} \left[\int_E p^\theta(x, x') \log \frac{p^\theta(x, x')}{p^{\theta+\epsilon}(x, x')} dx' \right].$$

(b) The proof of this part of the theorem is similar to the proof of the pathwise FIM for the relative entropy rate in [13]. We present it with minor but necessary adaptations. Let $\delta p(x, x') = p^{\theta+\epsilon}(x, x') - p^\theta(x, x')$, then the instantaneous relative entropy $\mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon})$ at the i -th time instant is written as

$$\begin{aligned} \mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon}) &= - \int_E \int_E \nu_{i-1}^\theta(x) p^\theta(x, x') \log \left(1 + \frac{\delta p(x, x')}{p^\theta(x, x')} \right) dx dx' \\ &= - \int_E \int_E \left[\nu_{i-1}^\theta(x) \delta p(x, x') - \frac{1}{2} \nu_{i-1}^\theta(x) \frac{\delta p(x, x')^2}{p^\theta(x, x')} + O(|\delta p(x, x')|^3) \right] dx dx'. \end{aligned}$$

Moreover, for all $x \in E$, it holds that

$$\int_E \delta p(x, x') dx' = \int_E p^{\theta+\epsilon}(x, x') dx' - \int_E p^\theta(x, x') dx' = 1 - 1 = 0.$$

Since the transition probability function is smooth w.r.t. the parameter vector θ , a Taylor series expansion to δp gives,

$$\delta p(x, x') = \epsilon^T \nabla_\theta p^\theta(x, x') + O(|\epsilon|^2).$$

Thus, we finally obtain for all $i = 1, \dots, T$, that

$$\begin{aligned}\mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon}) &= \frac{1}{2} \int_E \int_E \nu_{i-1}^\theta(x) \frac{(\epsilon^T \nabla_\theta p^\theta(x, x'))^2}{p^\theta(x, x')} dx dx' + O(|\epsilon|^3) \\ &= \frac{1}{2} \epsilon^T \left(\int_E \int_E \nu_{i-1}^\theta(x) p^\theta(x, x) \nabla_\theta \log p^\theta(x, x') \nabla_\theta \log p^\theta(x, x')^T dx dx' \right) \epsilon + O(|\epsilon|^3) \\ &= \frac{1}{2} \epsilon^T \mathcal{I}_\mathcal{H}(Q_i^\theta) \epsilon + O(|\epsilon|^3)\end{aligned}$$

where,

$$\mathcal{I}_\mathcal{H}(Q_i^\theta) = \mathbb{E}_{\nu_{i-1}^\theta} \left[\int_E p^\theta(x, x') \nabla_\theta \log p^\theta(x, x') \nabla_\theta \log p^\theta(x, x')^T dx' \right]$$

is the instantaneous FIM associated to the instantaneous relative entropy.

Consequently, the pathwise FIM $\mathcal{I}(Q_{0:T}^\theta)$, i.e., the Hessian of the pathwise relative entropy (9) at point θ , is given by

$$\mathcal{I}(Q_{0:T}^\theta) = \mathcal{I}(\nu^\theta) + \sum_{i=1}^T \mathcal{I}_\mathcal{H}(Q_i^\theta),$$

where $\mathcal{I}(\nu^\theta) = \mathbb{E}_{\nu^\theta}[\nabla_\theta \log \nu^\theta(x) \nabla_\theta \log \nu^\theta(x)^T]$ is the FIM of the initial distribution. \square

Remark 1: Let ‘ \otimes ’ denote the product operator of two distributions (i.e., $\nu \otimes p(A \times B) = \int_A p(x, B) \nu(x) dx$). Then, the instantaneous relative entropy can be written as a relative entropy of the probability measure $\nu_{i-1}^\theta \otimes p^\theta$ w.r.t. the probability measure $\nu_{i-1}^\theta \otimes p^{\theta+\epsilon}$. Mathematically,

$$\mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon}) = \mathcal{R}(\nu_{i-1}^\theta \otimes p^\theta | \nu_{i-1}^\theta \otimes p^{\theta+\epsilon}),$$

and similarly for the associated instantaneous FIM it holds that

$$\mathcal{I}_\mathcal{H}(Q_i^\theta) = \mathcal{I}(\nu_{i-1}^\theta \otimes p^\theta).$$

Remark 2: The instantaneous relative entropy $\mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon})$ is different from and should not be confused with the relative entropy of the unperturbed distribution at the i -th (or the $(i-1)$ -th) time instant ν_i^θ (or ν_{i-1}^θ) w.r.t. the respective perturbed distribution $\nu_i^{\theta+\epsilon}$ (or $\nu_{i-1}^{\theta+\epsilon}$). Indeed, it holds that

$$\mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon}) = \mathcal{R}(\nu_{i-1}^\theta \otimes p^\theta | \nu_{i-1}^\theta \otimes p^{\theta+\epsilon}) \neq \mathcal{R}(\nu_i^\theta | \nu_i^{\theta+\epsilon}),$$

as well as $\mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon}) \neq \mathcal{R}(\nu_{i-1}^\theta | \nu_{i-1}^{\theta+\epsilon})$. Moreover, an explicit formula for the probability distribution at the i -th time instant ν_i^θ is generally not available making the computation of the relative entropy $\mathcal{R}(\nu_i^\theta | \nu_i^{\theta+\epsilon})$ intractable. On the other hand, the instantaneous relative entropy $\mathcal{H}(Q_i^\theta | Q_i^{\theta+\epsilon})$ can be computed in a straightforward manner as a statistical average since it incorporates only the transition probabilities which are known functions.

Stationary regime: In the stationary regime, the initial distribution is the stationary distribution, $\mu^\theta(\cdot)$. Thus, for all $i = 1, \dots, T$ it holds that $\nu_i^\theta = \mu^\theta$ and the instantaneous relative entropy is a constant function of time that is equal to the relative entropy rate. The next corollary presents explicit formulas for the relative entropy rate and the associated FIM.

Corollary 3.1. *The relative entropy rate equals to*

$$\mathcal{H}(Q^\theta | Q^{\theta+\epsilon}) = \mathbb{E}_{\mu^\theta} \left[\int_E p^\theta(x, x') \log \frac{p^\theta(x, x')}{p^{\theta+\epsilon}(x, x')} dx' \right]. \quad (13)$$

Similarly, the FIM associated to the RER is given by

$$\mathcal{I}_\mathcal{H}(Q^\theta) = \mathbb{E}_{\mu^\theta} \left[\int_E p^\theta(x, x) \nabla_\theta \log p^\theta(x, x') \nabla_\theta \log p^\theta(x, x')^T dx' \right]. \quad (14)$$

Proof. Both formulas are obtained by substituting the stationary distribution μ^θ to the place of ν_{i-1}^θ in (10) and (12). \square

4 Continuous-time Markov chains

Let $\{X_t\}_{t \in \mathbb{R}_+}$ be a continuous-time Markov chain with countable state space E . The parameter dependent transition rates, denoted by $c^\theta(x, x')$, completely define the continuous-time Markov chain. The transition rates determine the updates (jumps or sojourn times) from a current state x to a new (random) state x' through the total rate $\lambda^\theta(x) = \sum_{x' \in E} c^\theta(x, x')$ which is the intensity of the exponential waiting time for a jump from state x . The transition probabilities for the embedded Markov chain $\{x_n\}_{n \geq 0}$ defined by $x_n := X_{t_n}$ where t_n is the instance of the n -th jump are $p^\theta(x, x') = \frac{c^\theta(x, x')}{\lambda^\theta(x)}$.

Assume another jump Markov process $\{\tilde{X}_t\}_{t \in \mathbb{R}_+}$, defined by perturbing the transition rates by a small vector $\epsilon \in \mathbb{R}^k$. Moreover assume that the two path probabilities $Q_{0:T}^\theta$ and $Q_{0:T}^{\theta+\epsilon}$ are absolutely continuous with respect to each other which is satisfied when $c^\theta(x, x') = 0$ if and only if $c^{\theta+\epsilon}(x, x') = 0, \forall x, x' \in E$. The following theorem presents the decomposition of the pathwise relative entropy for the case of continuous-time Markov chains.

Theorem 4.1. (a) *The pathwise relative entropy for the above-defined continuous-time Markov chain is decomposed as*

$$\mathcal{R}(Q_{0:T}^\theta | Q_{0:T}^{\theta+\epsilon}) = \mathcal{R}(\nu^\theta | \nu^{\theta+\epsilon}) + \int_0^T \mathcal{H}(Q_t^\theta | Q_t^{\theta+\epsilon}) dt, \quad (15)$$

where the instantaneous relative entropy equals to

$$\mathcal{H}(Q_t^\theta | Q_t^{\theta+\epsilon}) = \mathbb{E}_{Q_{0:t}^\theta} \left[\lambda^\theta(X_{t-}) \log \frac{c^\theta(X_{t-}, X_t)}{c^{\theta+\epsilon}(X_{t-}, X_t)} - (\lambda^\theta(X_t) - \lambda^{\theta+\epsilon}(X_t)) \right]. \quad (16)$$

(b) *Under smoothness assumption on the transition rate function $c^\theta(\cdot, \cdot)$ for the parameter θ , the pathwise FIM is also decomposed as*

$$\mathcal{I}(Q_{0:T}^\theta) = \mathcal{I}(\nu^\theta) + \int_0^T \mathcal{I}_{\mathcal{H}}(Q_t^\theta) dt, \quad (17)$$

where the instantaneous pathwise FIM is given by

$$\mathcal{I}_{\mathcal{H}}(Q_t^\theta) = \mathbb{E}_{Q_{0:t}^\theta} [\lambda^\theta(X_{t-}) \nabla_\theta \log c^\theta(X_{t-}, X_t) \nabla_\theta \log c^\theta(X_{t-}, X_t)^T]. \quad (18)$$

Proof. (a) As in the discrete-time case, the key element is an explicit formula for the Radon-Nikodym derivative. The Radon-Nikodym derivative of the path distribution $Q_{0:T}^\theta$ w.r.t. the perturbed path distribution $Q_{0:T}^{\theta+\epsilon}$ has an explicit formula known also as Girsanov formula [23, 18],

$$\frac{dQ_{0:T}^\theta}{dQ_{0:T}^{\theta+\epsilon}}(\{X_t\}_{t=0}^T) = \frac{\nu^\theta(X_0)}{\nu^{\theta+\epsilon}(X_0)} \exp \left\{ \int_0^T \log \frac{c^\theta(X_{t-}, X_t)}{c^{\theta+\epsilon}(X_{t-}, X_t)} dN_t - \int_0^T [\lambda^\theta(X_t) - \lambda^{\theta+\epsilon}(X_t)] dt \right\},$$

where ν^θ (reps. $\nu^{\theta+\epsilon}$) is the initial distributions of $\{X_t\}_{t \in \mathbb{R}_+}$ (resp. $\{\tilde{X}_t\}_{t \in \mathbb{R}_+}$) while N_t is the counting measure, i.e. counts the number of jumps in the process up to time t . Using

the Girsanov formula, the pathwise relative entropy is rewritten as

$$\begin{aligned}
& \mathcal{R}(Q_{0:T}^\theta | Q_{0:T}^{\theta+\epsilon}) \\
&= \mathbb{E}_{Q_{0:T}^\theta} \left[\log \frac{\nu^\theta(X_0)}{\nu^{\theta+\epsilon}(X_0)} + \int_0^T \log \frac{c^\theta(X_{t-}, X_t)}{c^{\theta+\epsilon}(X_{t-}, X_t)} dN_t - \int_0^T [\lambda^\theta(X_t) - \lambda^{\theta+\epsilon}(X_t)] dt \right] \\
&= \mathbb{E}_{Q_{0:T}^\theta} \left[\log \frac{\nu^\theta(X_0)}{\nu^{\theta+\epsilon}(X_0)} \right] + \mathbb{E}_{Q_{0:T}^\theta} \left[\int_0^T \log \frac{c^\theta(X_{t-}, X_t)}{c^{\theta+\epsilon}(X_{t-}, X_t)} dN_t \right] - \dots \\
&\quad \dots - \mathbb{E}_{Q_{0:T}^\theta} \left[\int_0^T [\lambda^\theta(X_t) - \lambda^{\theta+\epsilon}(X_t)] dt \right].
\end{aligned}$$

Exploiting the fact that the process $M_T := N_T - \int_0^T \lambda^\theta(X_t) dt$ is a martingale, we have that

$$\mathbb{E}_{Q_{0:T}^\theta} \left[\int_0^T \log \frac{c^\theta(X_{t-}, X_t)}{ac^\theta + \epsilon(X_{t-}, X_t)} dN_t \right] = \mathbb{E}_{Q_{0:T}^\theta} \left[\int_0^T \lambda^\theta(X_{t-}) \log \frac{c^\theta(X_{t-}, X_t)}{c^{\theta+\epsilon}(X_{t-}, X_t)} dt \right].$$

Thus, the pathwise relative entropy is rewritten as

$$\begin{aligned}
& \mathcal{R}(Q_{0:T}^\theta | Q_{0:T}^{\theta+\epsilon}) \\
&= \mathcal{R}(\nu^\theta | \nu^{\theta+\epsilon}) + \mathbb{E}_{Q_{0:T}^\theta} \left[\int_0^T \lambda^\theta(X_{t-}) \log \frac{c^\theta(X_{t-}, X_t)}{c^{\theta+\epsilon}(X_{t-}, X_t)} - (\lambda^\theta(X_t) - \lambda^{\theta+\epsilon}(X_t)) dt \right] \\
&= \mathcal{R}(\nu^\theta | \nu^{\theta+\epsilon}) + \int_0^T \mathbb{E}_{Q_{0:T}^\theta} \left[\lambda^\theta(X_{t-}) \log \frac{c^\theta(X_{t-}, X_t)}{c^{\theta+\epsilon}(X_{t-}, X_t)} - (\lambda^\theta(X_t) - \lambda^{\theta+\epsilon}(X_t)) \right] dt \\
&= \mathcal{R}(\nu^\theta | \nu^{\theta+\epsilon}) + \int_0^T \mathcal{H}(Q_t^\theta | Q_t^{\theta+\epsilon}) dt,
\end{aligned}$$

where the instantaneous relative entropy is defined as

$$\mathcal{H}(Q_t^\theta | Q_t^{\theta+\epsilon}) = \mathbb{E}_{Q_{0:t}^\theta} \left[\lambda^\theta(X_{t-}) \log \frac{c^\theta(X_{t-}, X_t)}{c^{\theta+\epsilon}(X_{t-}, X_t)} - (\lambda^\theta(X_t) - \lambda^{\theta+\epsilon}(X_t)) \right].$$

(b) Even though not directly evident from (16), the instantaneous relative entropy for continuous-time Markov chains is locally a quadratic function of the parameter vector θ . Indeed, defining the rate difference $\delta c(x, x') = c^{\theta+\epsilon}(x, x') - c^\theta(x, x')$, the instantaneous relative entropy can be rewritten as

$$\begin{aligned}
& \mathcal{H}(Q_t^\theta | Q_t^{\theta+\epsilon}) \\
&= -\mathbb{E}_{Q_{0:t}^\theta} \left[\lambda^\theta(X_{t-}) \log \left(1 + \frac{\delta c(X_{t-}, X_t)}{c^\theta(X_{t-}, X_t)} \right) \right] - \mathbb{E}_{Q_{0:t}^\theta} [\lambda^\theta(X_t) - \lambda^{\theta+\epsilon}(X_t)] \\
&= -\mathbb{E}_{Q_{0:t}^\theta} \left[\lambda^\theta(X_{t-}) \left(\frac{\delta c(X_{t-}, X_t)}{c^\theta(X_{t-}, X_t)} - \frac{1}{2} \left(\frac{\delta c(X_{t-}, X_t)}{c^\theta(X_{t-}, X_t)} \right)^2 + O(|\delta c(X_{t-}, X_t)|^3) \right) \right] \\
&\quad + \mathbb{E}_{Q_{0:t}^\theta} [\lambda^{\theta+\epsilon}(X_t) - \lambda^\theta(X_t)] \\
&= \frac{1}{2} \mathbb{E}_{Q_{0:t}^\theta} \left[\lambda^\theta(X_{t-}) \left(\frac{\delta c(X_{t-}, X_t)}{c^\theta(X_{t-}, X_t)} \right)^2 \right] + O(|\delta c|^3)
\end{aligned}$$

Under smoothness assumption on the transition rates in a neighborhood of parameter vector

θ a Taylor series expansion of $\delta c(x, x') = \epsilon^T \nabla_\theta c^\theta(x, x') + O(|\epsilon|^2)$ results in

$$\begin{aligned}
\mathcal{H}(Q_t^\theta | Q_t^{\theta+\epsilon}) &= \frac{1}{2} \mathbb{E}_{Q_{0:t}^\theta} \left[\lambda^\theta(X_{t-}) \left(\frac{\delta c(X_{t-}, X_t)}{c^\theta(X_{t-}, X_t)} \right)^2 \right] + O(|\delta c|^3) \\
&= \frac{1}{2} \epsilon^T \mathbb{E}_{Q_{0:t}^\theta} \left[\lambda^\theta(X_{t-}) \frac{\nabla_\theta c^\theta(X_{t-}, X_t) \nabla_\theta c^\theta(X_{t-}, X_t)^T}{c^\theta(X_{t-}, X_t)^2} \right] \epsilon + O(|\epsilon|^3) \\
&= \frac{1}{2} \epsilon^T \mathcal{I}_\mathcal{H}(Q_t^\theta) \epsilon + O(|\epsilon|^3)
\end{aligned}$$

where

$$\mathcal{I}_\mathcal{H}(Q_t^\theta) = \mathbb{E}_{Q_{0:t}^\theta} [\lambda^\theta(X_{t-}) \nabla_\theta \log c^\theta(X_{t-}, X_t) \nabla_\theta \log c^\theta(X_{t-}, X_t)^T] ,$$

is the instantaneous FIM. Finally, the pathwise FIM is obtained from a straightforward expansion of each element of the pathwise relative entropy in terms of ϵ . It is given by

$$\mathcal{I}(Q_{0:T}^\theta) = \mathcal{I}(\nu^\theta) + \int_0^T \mathcal{I}_\mathcal{H}(Q_t^\theta) dt ,$$

where $\mathcal{I}(\nu^\theta)$ is the FIM of the initial distribution while $\mathcal{I}_\mathcal{H}(Q_t^\theta)$ is the instantaneous pathwise FIM computed above. \square

Stationary regime: In the stationary regime, the instantaneous relative entropy is a constant function of time since at each time instant the distribution of the states is the –typically unknown– stationary distribution. The following corollary presents explicit formulas for the relative entropy rate and the associated FIM at the stationary regime.

Corollary 4.1. *The relative entropy rate is equal to the ergodic average*

$$\mathcal{H}(Q^\theta | Q^{\theta+\epsilon}) = \mathbb{E}_{\mu^\theta} \left[\sum_{x' \in E} c^\theta(x, x') \log \frac{c^\theta(x, x')}{c^{\theta+\epsilon}(x, x')} - (\lambda^\theta(x) - \lambda^{\theta+\epsilon}(x)) \right] , \quad (19)$$

while the FIM of the relative entropy rate, computed as its Hessian, has explicit formula given by

$$\mathcal{I}_\mathcal{H}(Q^\theta) = \mathbb{E}_{\mu^\theta} \left[\sum_{x' \in E} c^\theta(x, x') \nabla_\theta \log c^\theta(x, x') \nabla_\theta \log c^\theta(x, x')^T \right] . \quad (20)$$

Proof. At the stationary regime, the instantaneous relative entropy is rewritten as

$$\begin{aligned}
\mathcal{H}(Q_t^\theta | Q_t^{\theta+\epsilon}) &= \mathbb{E}_{Q_{0:t}^\theta} \left[\lambda^\theta(X_{t-}) \log \frac{c^\theta(X_{t-}, X_t)}{c^{\theta+\epsilon}(X_{t-}, X_t)} - (\lambda^\theta(X_t) - \lambda^{\theta+\epsilon}(X_t)) \right] \\
&= \mathbb{E}_{\mu^\theta} \left[\sum_{x' \in E} c^\theta(x, x') \log \frac{c^\theta(x, x')}{c^{\theta+\epsilon}(x, x')} - (\lambda^\theta(x) - \lambda^{\theta+\epsilon}(x)) \right] .
\end{aligned} \quad (21)$$

\square

5 Stochastic differential equations - Markov processes

Consider a Markov process $X_t \in \mathbb{R}^d$ driven by a stochastic differential equation of the form

$$\begin{cases} dX_t = b^\theta(X_t)dt + \sigma(X_t)dW_t \\ X_0 \sim \nu^\theta \end{cases} \quad (22)$$

where $b^\theta(\cdot)$ is the drift function depending on the parameter vector θ , $\sigma(\cdot)$ is the state-dependent diffusion matrix, W_t is a d -dimensional Brownian motion while ν^θ is the initial distribution of the process. Let $Q_{0:T}^\theta$ denote the path space distribution for a specific parameter vector θ . Consider also a perturbation vector, $\epsilon \in \mathbb{R}^K$, and denote by $Q_{0:T}^{\theta+\epsilon}$ the path space distribution of the perturbed process, \bar{X}_t driven by the stochastic differential equation (22) with perturbed drift function.

Under appropriate assumption on the components of the stochastic differential equation, the pathwise relative entropy can be decomposed as in the previous cases and an explicit formula for the instantaneous relative entropy can be estimated as the following theorem asserts.

Theorem 5.1. (a) Assume that the diffusion matrix, $\sigma(x)$, is invertible for all $x \in \mathbb{R}^d$ and

$$\mathbb{E}_{Q_{0:T}^\theta}[\exp \{ \int_0^T |\sigma^{-1}(X_t)(b^{\theta+\epsilon}(X_t) - b^\theta(X_t))|^2 dt \}] < \infty \quad (\text{Novikov condition}).$$

Then, the pathwise relative entropy for the above-defined Markov process is decomposed as

$$\mathcal{R}(Q_{0:T}^\theta | Q_{0:T}^{\theta+\epsilon}) = \mathcal{R}(\nu^\theta | \nu^{\theta+\epsilon}) + \int_0^T \mathcal{H}(Q_t^\theta | Q_t^{\theta+\epsilon}) dt, \quad (23)$$

where the instantaneous relative entropy is equal to

$$\mathcal{H}(Q_t^\theta | Q_t^{\theta+\epsilon}) = \frac{1}{2} \mathbb{E}_{\nu_t^\theta} \left[|\sigma^{-1}(x)(b^{\theta+\epsilon}(x) - b^\theta(x))|^2 \right]. \quad (24)$$

(b) Assume further that the drift function $b^\theta(\cdot)$ is smooth w.r.t. the parameter vector θ . Then, the pathwise FIM has a similar decomposition and the instantaneous FIM is given by

$$\mathcal{I}_{\mathcal{H}}(Q_t^\theta) = \mathbb{E}_{\nu_t^\theta} \left[[\nabla_\theta b^\theta(x)^T (\sigma(x)\sigma(x)^T)^{-1} \nabla_\theta b^\theta(x)] \right]. \quad (25)$$

where $\nabla_\theta b^\theta(\cdot)$ is a $d \times K$ matrix containing all the first-order partial derivatives of the drift vector (i.e., the Jacobian matrix).

Proof. (a) The inversion of the diffusion matrix is a necessary assumption for the well-posedness of the Novikov condition which in turn suffices for the two path distributions $Q_{0:T}^\theta$ and $Q_{0:T}^{\theta+\epsilon}$ to be absolutely continuous w.r.t. each other [24]. Additionally, the Girsanov theorem provides an explicit formula of the Radon-Nikodym derivative [24] which is given by

$$\frac{dQ_{0:T}^\theta}{dQ_{0:T}^{\theta+\epsilon}} \left(\{X_t\}_{t=0}^T \right) = \frac{d\nu^\theta}{d\nu^{\theta+\epsilon}}(X_0) \exp \left\{ - \int_0^T u(X_t)^T dW_t - \frac{1}{2} \int_0^T |u(X_t)|^2 dt \right\},$$

where $u(x) = \sigma^{-1}(x)(b^{\theta+\epsilon}(x) - b^\theta(x))$. Furthermore, it holds that

$$\hat{W}_t := \int_0^t u(X_s)ds + W_t,$$

is a Brownian motion w.r.t. the unperturbed path distribution $Q_{0:T}^\theta$, meaning that, for any measurable function $f(\cdot)$, it holds $\mathbb{E}_{Q_{0:T}^\theta} \left[\int_0^T f(X_t)^T d\hat{W}_t \right] = 0$. Then,

$$\begin{aligned}
\mathcal{R}(Q_{0:T}^\theta | Q_{0:T}^{\theta+\epsilon}) &= \mathbb{E}_{Q_{0:T}^\theta} \left[\log \frac{d\nu^\theta}{d\nu^{\theta+\epsilon}}(X_0) - \int_0^T u(X_t)^T dW_t - \frac{1}{2} \int_0^T |u(X_t)|^2 dt \right] \\
&= \mathbb{E}_{Q_{0:T}^\theta} \left[\log \frac{d\nu^\theta}{d\nu^{\theta+\epsilon}}(X_0) \right] - \mathbb{E}_{Q_{0:T}^\theta} \left[\int_0^T u(X_t)^T d\hat{W}_t \right] + \frac{1}{2} \mathbb{E}_{Q_{0:T}^\theta} \left[\int_0^T |u(X_t)|^2 dt \right] \\
&= \mathcal{R}(\nu^\theta | \nu^{\theta+\epsilon}) + \frac{1}{2} \mathbb{E}_{Q_{0:T}^\theta} \left[\int_0^T |u(X_t)|^2 dt \right] \\
&= \mathcal{R}(\nu^\theta | \nu^{\theta+\epsilon}) + \int_0^T \frac{1}{2} \mathbb{E}_{Q_{0:T}^\theta} [|u(X_t)|^2] dt \\
&= \mathcal{R}(\nu^\theta | \nu^{\theta+\epsilon}) + \int_0^T \frac{1}{2} \mathbb{E}_{\nu_t^\theta} |u(X_t)|^2 dt
\end{aligned}$$

hence the instantaneous relative entropy is explicitly given by

$$\mathcal{H}(Q_t^\theta | Q_t^{\theta+\epsilon}) = \frac{1}{2} \mathbb{E}_{\nu_t^\theta} [|\sigma^{-1}(x)(b^{\theta+\epsilon}(x) - b^\theta(x))|^2]$$

(b) Due to the smoothness assumption, a Taylor expansion of the drift function around the point θ results in $b^{\theta+\epsilon}(x) - b^\theta(x) = \nabla_\theta b^\theta(x)\epsilon + O(|\epsilon|^2)$ where $\nabla_\theta b^\theta(\cdot)$ is a $d \times K$ matrix containing all the first-order partial derivatives of the drift vector function (i.e., the Jacobian matrix). Then, it is straightforward to obtain from (24) to get

$$\mathcal{I}_\mathcal{H}(Q_t^\theta) = \mathbb{E}_{\nu_t^\theta} [\nabla_\theta b^\theta(x)^T (\sigma(x)\sigma(x)^T)^{-1} \nabla_\theta b^\theta(x)] .$$

□

Stationary regime: In the stationary regime, the instantaneous relative entropy becomes a constant function of time since the distribution of the process equals to the stationary distribution denoted by μ^θ for all times. The following corollary presents explicit formulas for the relative entropy rate and the associated FIM.

Corollary 5.1. *Under the same assumption of Theorem 5.1 and let $X_0 \sim \mu^\theta$. Then, the relative entropy rate equals to*

$$\mathcal{R}(Q^\theta | Q^{\theta+\epsilon}) = \frac{1}{2} \mathbb{E}_{\mu^\theta} [|\sigma^{-1}(x)(b^{\theta+\epsilon}(x) - b^\theta(x))|^2] , \quad (26)$$

while the FIM associated with the relative entropy rate is given by

$$\mathcal{I}_\mathcal{H}(Q^\theta) = \mathbb{E}_{\mu^\theta} [\nabla_\theta b^\theta(x)^T (\sigma(x)\sigma(x)^T)^{-1} \nabla_\theta b^\theta(x)] . \quad (27)$$

We finally remark that a popular method for modeling non-equilibrium systems in atomistic and mesoscopic scales is based on the Langevin equation. Langevin equation is a degenerate system of stochastic differential equations whose sensitivity analysis based on the relative entropy rate and the associated pathwise FIM was performed in [25].

6 Demonstration example

In this section we give a numerical example of the pathwise relative entropy (6) and pathwise FIM (8) for a continuous time Markov chain model. More specific, a biological reaction network is considered and the quantities instantaneous RE (16) and instantaneous FIM (18) are presented as a function of time.

6.1 Continuous time Markov chains: an EGFR model

In [26], Kholodenko et al. proposed a reaction network that describes signaling phenomena of mammalian cells [27, 28, 29]. The reaction network consists of $N = 23$ species and $M = 47$ reactions. The propensity function for the R_j reaction, $j = 1, \dots, 47$ and $j \neq 7, 14, 29$, obeys the law of mass action [2],

$$a_j(\mathbf{x}) = k_j \binom{\mathbf{x}_{A_j}}{\alpha_j} \binom{\mathbf{x}_{B_j}}{\beta_j}, \quad (28)$$

for a reaction of the general form “ $\alpha_j A_j + \beta_j B_j \xrightarrow{k_j} \dots$ ”, where A_j and B_j are the reactant species, α_j and β_j are the respective number of molecules needed for the reaction, k_j the reaction constant and \mathbf{x}_{A_j} and \mathbf{x}_{B_j} is the total number of species A_j and B_j , respectively. The binomial coefficient is defined by $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. The propensity functions for reactions R_7, R_{14}, R_{29} are being described by the Michaelis–Menten kinetics, see [2],

$$a_j(\mathbf{x}) = V_j \mathbf{x}_{A_j} / (K_j + \mathbf{x}_{A_j}), \quad j = 7, 14, 29, \quad (29)$$

where V_j represents the maximum rate achieved by the system at maximum (saturating) substrate concentrations while K_j is the substrate concentration at which the reaction rate is half the maximum value. The parameter vector contains all the reaction constants,

$$\theta = [k_1, \dots, k_6, k_8, \dots, k_{13}, k_{15}, \dots, k_{28}, k_{30}, \dots, k_{47}, V_7, K_7, V_{14}, K_{14}, V_{29}, K_{29}]^T. \quad (30)$$

In this study the values of the reaction constants are the same as in [26].

For the initial data and parameters chosen in this study, the time series can be split into two regimes: (a) a transient regime that approximately corresponds to the time interval $[0, 50]$ and (b) a stationary regime which approximately corresponds to the time interval $[50, \infty)$.

Next, we discuss two sensitivity measures of the process $\{X_s\}_{s=0}^t$: the instantaneous RE defined in (15)

$$\begin{aligned} f(\{X_s\}_{s=0}^t) &= \mathcal{H}(Q_t^\theta | Q_t^{\theta+\epsilon}) \\ &= \mathbb{E}_{\nu_t^\theta} \left[\sum_{j=1}^M a_j^\theta(\mathbf{x}) \log \frac{a_j^\theta(\mathbf{x})}{a_j^{\theta+\epsilon}(\mathbf{x})} - (a_0^\theta(\mathbf{x}) - a_0^{\theta+\epsilon}(\mathbf{x})) \right], \end{aligned} \quad (31)$$

with $t \in [0, T]$ and the averaged RE, defined as

$$g(\{X_s\}_{s=0}^t) = \frac{1}{t} \int_0^t \mathcal{H}(Q_s^\theta | Q_s^{\theta+\epsilon}) ds, \quad t \in [0, T]. \quad (32)$$

In Figure (1) the two sensitivity measures are presented for $T = 100$. As expected, the averaged RE is smoother than the instantaneous RE while some of the qualitative characteristics remain. On the other hand, quantitative characteristics, such as the time that two instantaneous RE are crossed, are not preserved in the averaged RE. The averaged RE in the interval $[0, t]$ should be interpreted as a measure of the information accumulated in the whole interval while the instantaneous RE is a measure of the information at the time instant t . Moreover, the observable can be used as part of an upper bound for a different sensitivity measure, see [15] for a detailed discussion.

Let us define a different sensitivity measure as the relative difference between the k -th species of two systems where the ℓ -th parameter of the second is perturbed by ϵ' ,

$$S_{k,\ell,t} := \frac{X_{k,t}^\theta - X_{k,t}^{\theta+\epsilon_\ell}}{X_{k,t}^\theta}. \quad (33)$$

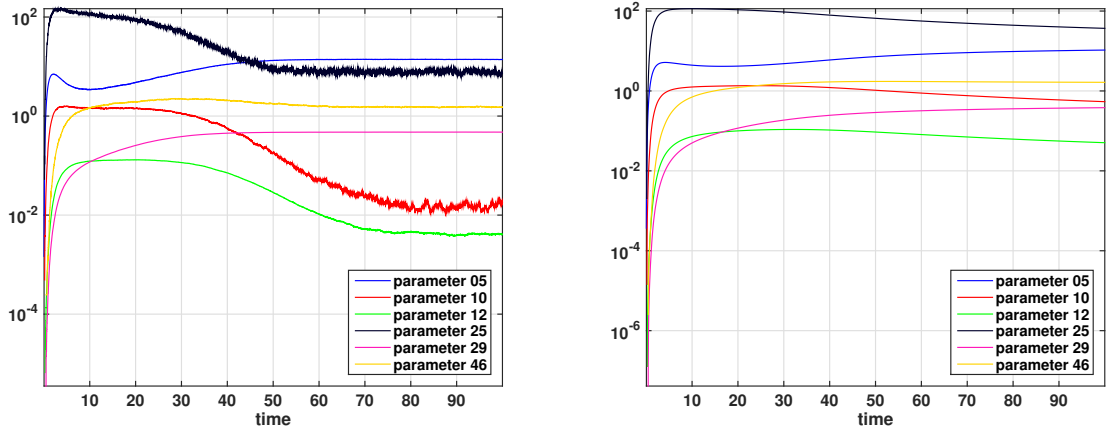


Figure 1: The instantaneous RE, defined in (15), for the EGFR model (left) and the averaged RE, defined in (32) (right).

where ϵ_ℓ is a vector with zeros everywhere and ϵ' in the ℓ -th position. In the following examples the value of ϵ' for a perturbation in the ℓ -th parameter is $\epsilon' = 0.1\theta_\ell$. By summing over all species we obtain a total sensitivity measure, which is only indexed by the parameter index and time,

$$S_{\ell,t} = \frac{1}{N} \sum_{k=1}^N S_{k,\ell,t} . \quad (34)$$

By observing the first row of Figure 2, which shows the instantaneous RE (16) for $\ell = 10$ and $\ell = 29$, we learn that perturbations in the 10-th parameter have large sensitivity for small times and as time varies the sensitivity is getting smaller. On the other hand, perturbations in the 29-th parameter have small influence on the system for small times while for larger times the sensitivity becomes significant. These observations are in good agreement with the sensitivity measure (35) presented in the second and third row of Figure 2.

Moreover there is a crossing in the instantaneous RE which happens around $t = 42$. After this time the sensitivity in parameter $\ell = 29$ becomes more significant than the sensitivity in parameter $\ell = 10$. This is again in agreement with the behavior of the sensitivity measure (35). This observation shows that the transient regime is more sensitive in perturbations in the 10-th parameters while the equilibrium is more sensitive in perturbations in the 29-th parameter.

7 Conclusions

In this paper we presented two pathwise sensitivity measures for the analysis of stochastic systems; the instantaneous relative entropy and its approximation, the instantaneous Fisher information matrix. These sensitivity tools serve as an extension to transient processes of the sensitivity tools presented in [13]. Three examples, discrete-time Markov chains, continuous-time Markov Chains and stochastic differential equations, were presented as an application of the new sensitivity measures. In Section 6 we demonstrated, in a biological reaction network, how the proposed pathwise sensitivity measure can be applied to transient, as well as in steady state regimes.

Finally, the pathwise sensitivity method is directly connected to a different sensitivity

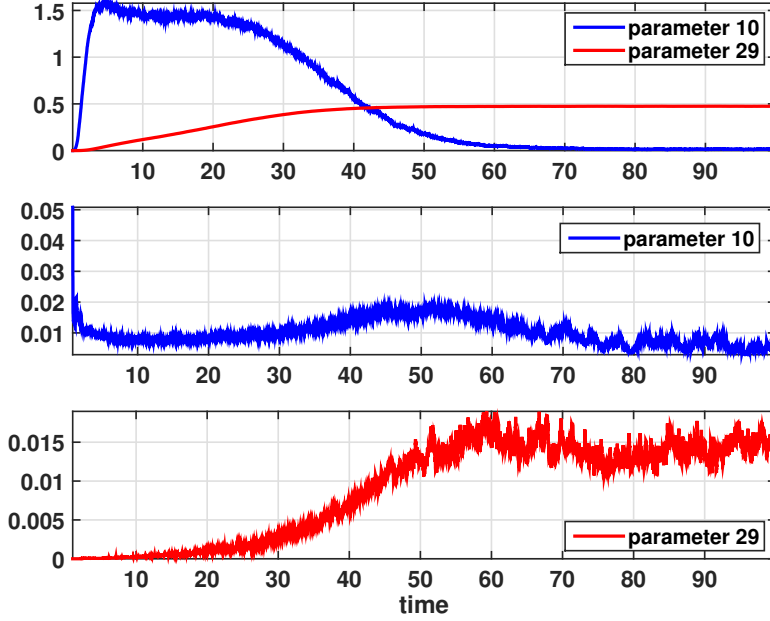


Figure 2: Instantaneous RE (15) for the EGFR model and parameters 10 and 29 (first row). The total sensitivity of the system, as defined in (34), due to perturbations in the 10-th parameter (second row) and due to the 29-th parameter (third row).

measure that depends on specific observables. More specifically, if we define the sensitivity index (SI) of the ℓ -th observable to the k -th parameter as

$$S_{k,\ell} = \frac{\partial}{\partial \theta_k} \mathbb{E} [f_\ell (\{X_s\}_{s=0}^T)] , \quad (35)$$

then IFIM (18) serves as an upper bound of $S_{k,\ell}$ through the inequality,

$$|S_{k,\ell}| \leq \sqrt{\text{Var}_{Q_{0:T}^\theta}(f_\ell)} \sqrt{\mathcal{I}(Q_{0:T}^\theta)_{k,k}} , \quad (36)$$

where $F = (f_1, \dots, f_L)$ is a vector of observable functions. This inequality follows by rearranging the generalized Cramer-Rao bound for a biased estimator [30, 31]. Due to low variance of the estimator of IFIM compared to the variance of a finite difference estimator of $S_{k,\ell}$, the estimation of the right hand side of (36) is faster than that of the left hand side of (36). In [15] the authors use this inequality to efficiently screen out and exclude low sensitivity indices under a pre-specified value and then perform a coupling finite difference algorithm [8] to accurately estimate the remaining sensitivity indices $S_{k,\ell}$. In Figure 3 the estimated SIs for the EGFR model discussed in Section 6 are ordered in the parameter direction using only the IFIM (18). Notice that the SIs are then grouped into four distinct regions. For a detailed presentation of this methodology we refer to [15].

Acknowledgement

The work of the authors was supported by the Office of Advanced Scientific Computing Research, U.S. Department of Energy, under Contract No. de-sc0002339 and by the Euro-

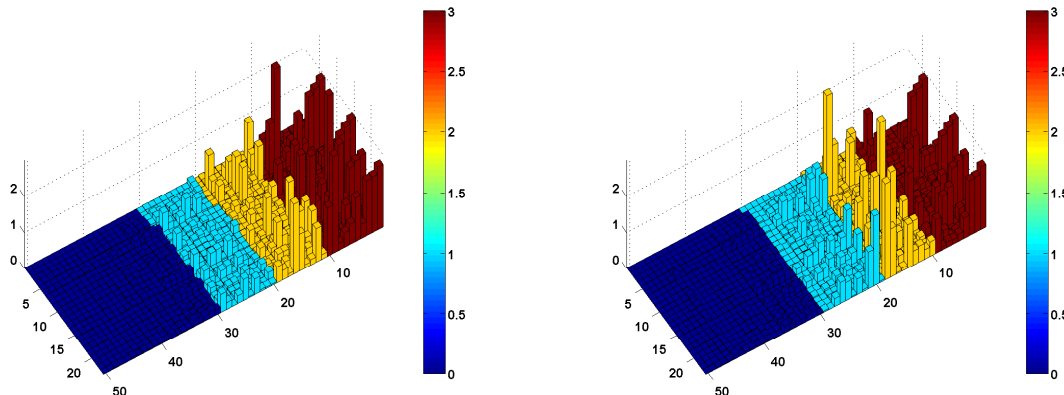


Figure 3: Ordering of the sensitivity index (35) in the time interval $[0, 50]$ (left) and $[0, 100]$ (right) utilizing the averaged IRE (32). In this case $f_\ell(\{X_s\}_{s=0}^T) = \frac{1}{T} \int_0^T X_{\ell,s} ds$, which is the mean concentration of the ℓ -th species in time interval $[0, T]$.

pean Union (European Social Fund) and Greece (National Strategic Reference Framework), under the THALES Program, grant AMOSICSS.

References

- [1] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis. The Primer*. Wiley, 2008.
- [2] J. DiStefano III. *Dynamic Systems Biology Modeling and Simulation*. Elsevier, 2013.
- [3] P.W. Glynn. Likelihood ratio gradient estimation for stochastic systems. *Communications of the ACM*, 33(10):75–84, 1990.
- [4] M. Nakayama, A. Goyal, and P. W. Glynn. Likelihood ratio sensitivity analysis for Markovian models of highly dependable systems. *Stochastic Models*, 10:701–717, 1994.
- [5] S. Piyasunov and A. P. Arkin. Efficient stochastic sensitivity analysis of discrete event systems. *J. Comp. Phys.*, 221:724–738, 2007.
- [6] D. Kim, B.J. Debusschere, and H.N. Najm. Spectral methods for parametric sensitivity in stochastic dynamical systems. *Biophysical Journal*, 92:379–393, 2007.
- [7] M. Rathinam, P. W. Sheppard, and M. Khammash. Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks. *J. Chem. Phys.*, 132:034103–(1–13), 2010.
- [8] David F. Anderson. An efficient finite difference method for parameter sensitivities of continuous-time Markov chains. *SIAM J. Numerical Analysis*, 50(5):2237–2258, 2012.
- [9] P.W. Sheppard, M. Rathinam, and M. Khammash. A pathwise derivative approach to the computation of parameter sensitivities in discrete stochastic chemical systems. *J. Chem. Phys.*, 136(3):034115, 2012.
- [10] H. Meskine, S. Matera, M. Scheffler, K. Reuter, and H. Metiu. Examination of the concept of degree of rate control by first-principles kinetic Monte Carlo simulations. *Surf. Science*, 603(10-12):1724–1730, 2009.

- [11] C. Maes M. Baiesi and B. Wynants. Nonequilibrium linear response for markov dynamics, i: jump processes and overdamped diffusions. *J. Stat. Phys.*, 2009.
- [12] E. Boksenbojm M. Baiesi, C. Maes and B. Wynants. Nonequilibrium linear response for markov dynamics, ii: Inertial dynamics. *J. Stat. Phys.*, 2010.
- [13] Y. Pantazis and M. Katsoulakis. A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics. *J. Chem. Phys.*, 138(5):054115, 2013.
- [14] P. Dupuis, M.A. Katsoulakis, Y. Pantazis, and P. Plecháč. Sensitivity bounds and error estimates for stochastic models. (*in preparation*).
- [15] G. Arampatzis and Y. Pantazis M. A. Katsoulakis. Accelerated sensitivity analysis in high-dimensional stochastic reaction networks. *Submitted to PLoS ONE*.
- [16] S. Kullback. *Information theory and statistics*. John Wiley and Sons, NY, 1959.
- [17] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.
- [18] C. Kipnis and C. Landim. *Scaling Limits of Interacting Particle Systems*. Springer-Verlag, 1999.
- [19] M. Avellaneda, C.A. Friedman, R. Holmes, and D.J. Samperi. Calibrating volatility surfaces via relative-entropy minimization. *Social Science Research Network*, 1997.
- [20] HB Liu, W Chen, and A Sudjianto. Relative entropy based method for probabilistic sensitivity analysis in engineering design. *J. Mechanical Design*, 128:326–336, 2006.
- [21] N. Limnios and G. Oprisan. *Semi-Markov Processes and Reliability*. Springer, 2001.
- [22] R. V. Abramov, M. J. Grote, and A. J. Majda. *Information Theory and Stochastics for Multiscale Nonlinear Systems*. CRM Monograph Series, 2005.
- [23] R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes: I & II*. Springer, 1977.
- [24] B Oksendal. *Stochastic Differential Equations: An introduction with applications*. Springer-Verlag, 2000.
- [25] A. Tsourtis, Y. Pantazis, and V. Harmandaris M.A. Katsoulakis. Parametric sensitivity analysis for stochastic molecular systems using information theoretic metrics. *Submitted to J of Chemical Physics*.
- [26] Boris N Kholodenko, Oleg V Demin, Gisela Moehren, and Jan B Hoek. Quantification of short term signaling by the epidermal growth factor receptor. *Journal of Biological Chemistry*, 274(42):30169–30181, 1999.
- [27] N. Moghal and P.W. Sternberg. Multiple positive and negative regulators of signaling by the egf receptor. *Curr. Opin. Cell. Biol.*, 11:190–196, 1999.
- [28] P.O. Hackel, E. Zwick, N. Prenzel, and A. Ullrich. Epidermal growth factor receptors: critical mediators of multiple receptor pathways. *Curr. Opin. Cell. Biol.*, 11:184–189, 1999.
- [29] B. Schoeberl, C. Eichler-Jonsson, E.D. Gilles, and G. Muller. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature Biotechnology*, 20:370–375, 2002.
- [30] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002.
- [31] S. M. Kay. *Funtamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Englewood Cliffs, NJ, 1993.